DOCUMENT RESUME

ED 355 383 CE 063 220

AUTHOR Lowry, Phillip E.
TITLE The Assessment Center: Do Assessor Characteristics Affect Scores?
PUB DATE Jun 92
NOTE 17p.; Supported in part by a grant from First Interstate Bank of Nevada.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Administrator Evaluation; Administrator Selection; Adults; *Assessment Centers (Personnel); Credentials; Evaluation Criteria; *Evaluation Methods; Evaluation Research; *Evaluators; *Experimenter Characteristics; Fire Fighters; *Interrater Reliability; Local Government; *Personnel Evaluation; Personnel Selection; Police; Situational Tests

ABSTRACT

The effects of 9 assessor characteristics on scores given by 49 assessors to 111 assessees in 3 police and 6 fire service assessment centers for local governments were investigated. Assessors included those who had and who did not have experience in the job under consideration, and who differed on several other characteristics, including the following: age, race, gender, previous assessment center experience, education, managerial experience, assessor managerial level, and tenure in that position. Age and the rank of the assessor were the only characteristics that had a significant effect on the scores. However, the magnitude of the effect was quite small (less than two percent of the variance was explained by these two factors). The results might be attributed to the process used to select the assessors, the way the centers were conducted, and the type and intensity of assessor training. The study concluded that assessor characteristics were minor factors in the overall scoring model. Reasons for this independence of scores from assessor characteristics may have included the following: (1) the model used for selecting the assessors, (2) the training provided for the assessors, (3) the use of highly exercise-specific behaviorally anchored rating forms, and (4) the use of an integrating discussion technique after each exercise in eight of the nine centers. (The report includes an appendix that lists criteria for selecting assessors). (Contains 15 references.) (Author/KC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED355383

The Assessment Center: Do Assessor Characteristics Affect Scores?

Phillip E. Lowry

Abstract

The effects of nine assessor characteristics on exercise scores given by 49 assessors to 111 assessees in three police and six fire service assessment centers for local governments were investigated. Assessors included those who had and who did not have experience in the job under consideration, and who differed on several other characteristics including: age, race, gender, previous assessment center experience, education, managerial experience, assessor managerial level, and tenure in that position. Age and the rank of the assessor were the only characteristics that had a significant effect on the scores. However, the magnitude of the effect was quite small (less than 2% of the variance was explained by these two factors). The results might be attributed to the process used to select the assessors, the way the centers were conducted, and the type and intensity of assessor training. Implications for practitioners and researchers are discussed.

The assessment center is an increasingly important tool for personnel selection in local governments. Fitzgerald and Quintance (1982) reported that over 44 percent of 156 federal, state, and local governments used the assessment center. As of 1985, 32 of 73 metropolitan fire departments used the assessment center, especially for promotion (Yeager, 1986). The primary reason for the continued acceptance of the assessment center method appears to be a relatively large literature that shows a positive relationship between assessment center scores and performance as a manager or supervisor (Thornton & Byham, 1982).

The assessment center process requires the use of assessors to observe and evaluate assessee behaviors. Clearly the assessor is a key factor in the process. There is, however, little published research on which, if any, assessor characteristics have a significant or important effect on scores. This paper addresses that issue.

Preparation of this paper was supported in part by a grant from the First Interstate Bank of Nevada. I wish to thank Professors John Schibrowsky and Thomas Carroll of the University of Nevada, Las Vegas, for their assistance in the statistical analyses used in this paper.

Requests for reprints should be sent to Phillip E. Lowry, Chair, Department of Public Administration, University of Nevada, Las Vegas, Las Vegas, NV 89154.

CE 063220

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

P Lowry

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Assessment Center Process

In a typical assessment center, a job analysis is used to identify specific tasks that must be performed properly for satisfactory service in the job under consideration. The knowledge, skills, and abilities (KSA) required to perform the tasks may also be identified. These KSAs are often referred to as "performance dimensions." One or more simulations are developed that test the tasks or performance dimensions. Assessors observe the behaviors of the participants and either determine a score on how well the task was performed, or on the selected performance dimension.

The current guidelines for assessment center operation were prepared by the Task Force on Assessment Center Guidelines and published in 1989 (Task Force, 1989). The Guidelines include relatively general guidance on the assessment center process and specific direction concerning the training goals for assessors (Task Force, 1989, pp. 465-468). The Guidelines, are however, silent on the specific characteristics that should be possessed by assessors.

Assessor Characteristics

This paper addresses whether assessor characteristics have a significant and important effect on the scores given by the assessor. The Guidelines include the following:

> 6. Multiple assessors must be used for each assessee.
>
> When selecting a group of assessors the following characteristics should be considered: diversity of ethnicity, age, gender, and functional work area.(Task Force, 1989, p. 462).

The Guidelines also outline the training goals for assessors and give some suggestions concerning how these goals should be accomplished (Task Force, 1989, pp. 463, 565-468).

Thornton and Byham (1982) suggest that in addition to thorough training in the procedures and exercises to be used in the center, that:

> The assessors should know what it takes to succeed. They should be thoroughly familiar with the job and the organization and, if possible, have experience in the job." (P.223).

Typically, assessors for police and fire service assessment centers are selected from similar agencies in other local governments. This is done to ensure the assessor is familiar with the job under test, and to minimize the chance that an assessor is intimately familiar with an assessee.

In many cases, local governments will also select assessors who are from the local community, but who are not affiliated with the profession under consideration. Their reasons for using local citizens typically have included a desire to ensure that representatives from the local citizenry at large, or persons from other departments within the local government have an opportunity to be involved in and lend their judgment and expertise to the promotion

process.

This practice has raised questions concerning the ability of persons who are not affiliated with the profession under consideration to judge assessees for promotion in that profession. In other words, is it reasonable to expect the nonaffiliated assessor, who has had no experience in the job under consideration, to evaluate assessees and give valid scores?

The professional affiliation of the assessor is not the only characteristic of interest. As the Guidelines suggest, age, gender, and ethnicity should also be considered. The primary concern about this issue is whether it is possible that the characteristics of the assessors could, in and of themselves, have a significant or important effect on the scores. The Guidelines imply that this might be the case; hence, consideration must be given to the issue. The question is, therefore, how much consideration? How important is it to ensure homogeneity of assessor characteristics?

There is little published research on the impact of assessor characteristics on scoring. The concerns raised about training of assessors, and their familiarity with the job are logical, but there have thus far been few reports of empirical tests of these concerns. There has been some published research on the training of assessors and the use of trained psychologists who are not necessarily familiar with either the organization or the job under consideration.

Gaugler, et al. (1987) did a meta-analysis of assessment centers and reported that the type of assessor, manager or psychologist, was the only variable, among several, that moderated the validities in the centers they analyzed. The variables that did not significantly moderate validity included: amount of assessor training, number of days spent observing, number of hours spent integrating information, the ratio of number of assessees to assessors, and to whom feedback was given. They went on to report that psychologists provided more valid ratings than in-house managers (p. 505).

Carrier, et al. (1990), in a selection interview situation, attempted to determine if there were differences in ratings among judges with differing levels of managerial expertise, or experience in selection. They found no significant differences in scores based on either of these characteristics.

Research Question and Purpose

This paper reports the results of analyzing nine assessment centers conducted by local governments to select fire and police professionals for promotion or hire (Table 1). The assessors were selected from among professional managers within the assessees' profession, i.e., affiliated with the profession, but outside the organization; and from managers who were from the local community but who were not affiliated with the assessees' profession.

In each center the assessors observed the behaviors of the

assessees performing a series of job related tasks. In all but one center, an integrating discussion was held after each exercise. The final score was based solely on the performance of the tasks in each exercise. I refer to this type of center as a "Task-Specific" center.

In Center 1, conducted under the previous standards (Task Force, 1980), the assessors recorded their scores after each exercise with no integrating discussion. Upon completion of all the exercises, the integrating discussion was conducted. The final score for each of the several performance dimensions was given by the assessors based on their observations and the results of the integrating discussion. I refer to this type of center as a "Dimension-Specific" center.

Research Question. The analyses to be reported here address the following research question: Are there differences in the scores given by the assessors based on their characteristics? The characteristics tested included AGE, GENDER, RACE, RANK/position (senior or middle level management), TENURE in that rank or position, MANagerial experience (number of years), ASSESSOR experience in other assessment centers, EDucation (years of formal education), and PROfessional affiliation (yes or no).

Purpose. The major purpose of this study was to assist those who must select assessors. The answers to the research question were sought to provide guidance concerning which, if any, characteristics ought to be the same for each assessor. The main concern was to determine if there is a significant risk to the validity of a center if the assessors differ on one or more characteristics.

Method

Data for this study were collected during nine selection assessment centers conducted for local governments. Details of the centers are in Table 1.

Table 1
The Assessment Centers

Center	Type	Rank Tested	Candidates	Assessors
1	Fire	Battalion Chief	10	4
2	Police	Captain	7	6
3	Police	Lieutenant	10	6
4	Fire	Captain	6	4
5	Fire	Battalion Chief	9	4
6	Police	Captain	8	4
7	Fire	Fire Chief	6	4
8	Fire	Captain	31	8
9	Fire	Captain	24	9
		Total	111	16

Assessees

The assessees in all but Center 7 were members of the organization conducting the center. The assessees in Center 7 were professional fire service officers from organizations located throughout the United States.

All the assessees were seeking promotion or appointment to a supervisory or managerial position.

Assessor selection

The assessors were selected by the sponsoring local government organization. All assessors were volunteers. Assessors from fire or police departments implicitly agreed to participate based on the assumption that they could call upon the sponsoring organization for assessors when they needed them. The other assessors were either local citizens or members of other local government departments. No direct compensation was offered any assessor; their expenses were paid by the sponsoring organization.

The sponsoring organizations were asked to select the assessors using the criteria listed in the Appendix. The assessors met these criteria with very few exceptions. Those exceptions were generally limited to the amount of experience and training as assessors. For example, in Center 7 the two nonaffiliated assessors had no previous assessment center experience or training in the process. By contrast, the affiliated assessors in that center had participated in several centers and had previously received at least five days of training.

The affiliated assessors were from organizations generally in the same state with some exceptions in Centers 8 and 9. None were from the organization conducting the center. In each center, the affiliated assessors were at or above the rank of the job under

consideration. The nonaffiliated assessors were either local citizens or members of other local government departments.

In all cases, the assessors were, or had been, supervisors or managers, primarily in local government or federal agencies. No assessor had ever worked with the assessees, or even knew the assessees with one exception. In that case, the two nonaffiliated assessors in Center 7 knew an assessee casually, but assured me that they could judge his performance objectively.

Table 2 summarizes the values of the measured characteristics of the assessors.

Table 2

Assessor Characteristics

Professional		Gender		Race			Experience[4]		Other[4]			
Yes	No[1]	M	F[2]	W	B	H[3]	Managerial	Assessor	Age	Ed	Rank	Tenure
33	16	46	3	38	5	6	5+yrs	2+ centers	36-45	16 yrs	Senior Mgt	3-5yrs

Notes [1]: Affiliated with profession; [2]: Male, Female; [3]: White, Black, Hispanic; [4]: Medians.

Preliminary Procedures

Job Analysis. For each assessment center, I either did a job analysis or used one that had been done within the past three years. The most important tasks required for entry to the job under consideration were identified. In Center 1, the dimension-specific center, essential performance dimensions were identified. These were leadership, written communications, oral communications, problem solving, and planning and organizing.

Exercises. Based on the job analysis, three or four simulations were prepared for each center. They were designed to allow the participants to perform the identified job-related tasks to demonstrate their competence either in the task, or in the identified performance dimensions.

In two of the fire captain centers, Centers 8 and 9, a fire scene simulation was used. In these and the other centers, other exercises included one or more of the following: a written analysis of two or more critical events, a written/oral analysis of a problem, a role playing exercise involving a personnel problem, and either a leaderless group or leader group discussion.

Assessor Training. The training each assessor received was the same in each center. The assessors were trained in observation and

evaluation of behaviors and in the specific exercises that would be used. Each was given a comprehensive training manual approximately one week before the center. The manual described the general rules that would be followed and gave specific instructions for observing, recording, and evaluating behaviors. It is estimated that the assessors spent an average of four hours reviewing the training manual.

During the training session conducted on the day prior to the first day of exercises, the assessors acted as both assessors and assessees using the same exercises that were used in the center. Their performance was closely observed and reviewed to ensure they were all familiar with the key elements of each exercise, and that they could perform their tasks effectively.

The assessors were encouraged to discuss freely the participant and assessor instructions, and the rating forms provided for each exercise. If they had questions concerning any aspect of the department's policies, procedures, relevant law, technical issues, or other matters of concern, they were discussed and answered.

Conduct of the Centers

The process used in the centers was based on the methodology suggested by Lowry (1991).

Assessors were randomly assigned to teams with one nonaffiliated assessor teamed with one or more affiliated assessors. The actual process for evaluation differed between Center 1 and the other centers.

Center 1, the dimension-specific center, was conducted under the previous standards for conducting assessment centers (Task Force, 1979). In this center, the assessors recorded their observations and privately recorded an exercise score after each exercise. After all the exercises were completed, they participated in an integrating discussion where they related their observations of important behaviors to the other assessors. After the discussion, they privately assigned a score for each of the performance dimensions. The integrating discussions did not include any evaluative comments nor were scores divulged. The final score for that performance dimension was either the arithmetic average of all the assessor's scores for that dimension, or the mode, as suggested by Sackett & Wilson (1982)[1]. The overall score for the center was the arithmetic average of the scores for each performance dimension.

The remaining centers were all task-specific centers conducted under the current Guidelines (Task Force, 1989). The assessors recorded their observations and after each exercise engaged in an

---

1. With four assessors, if there is more than a one scale difference among the scores, use the mode. In all other cases, use the mean.

integrating discussion of these behaviors. The integrating discussion is optional under the current Guidelines. It was used because I have determined from discussions with assessors in previous assessment centers that they want and need to hear what the other assessors observed.

Again, no evaluative comments were allowed, and scores were not divulged. Upon completion of the discussions, the assessors recorded the final score for the exercise. The total score for the exercise was also determined using the Sackett & Wilson (1982) technique. The total overall score was the arithmetic average of all exercise scores.

A behaviorally anchored observation and rating form was used in each exercise. This form included specific issues considered essential for proper performance in the exercise. It also included the rating scale with definitions for each of the scale values. Reilly, et al. (1990) suggest that such checklists reduce "the cognitive demands placed on raters."(p.71).

The rating scale in all centers except Center 8 was a five point scale. The scale used in Center 8 was an eight point scale. The assessors were required to use an integer score.

One experienced assessor who received special training was appointed a facilitator. The duties of the facilitator included the conduct of the integrating discussions. The facilitator ensured that the rules governing the integrating discussion, especially the restriction on evaluative discussions and divulging of ratings, were followed.

Results

The database used for this research included the assessor characteristics and scores given in each exercise in each center to the assessees. There was a total of 111 assessees and 49 assessors. The n in this database was 1299 cases.

SPSSPC+ was used for all the data analyses in this research. The first step in the analysis was to examine the product moment correlations to determine if there were any significant relationships among the assessor characteristics (Table 3). There is considerable intercorrelation among the assessor characteristics. For example, all the characteristics are significantly related to PRO, the professional affiliation of the assessor. Therefore, in order to answer the research question, it was first necessary to remove the effects of the intercorrelations.

Table 3

Intercorrelations of Variables

Variable	1	2	3	4	5	6	7	8	9
1. SCORE									
2. AGE	.14								
3. SEX	.04	.20							
4. RACE	.02	.01	.16						
5. RANK	.07	.31	.28	.18					
6. TENURE	.07	.49	.19	-.09	.08				
7. MAN	.01	.18	.47	.15	.14	.36			
8. ASSESSOR	.01	.05	-.08	.03	.00	-.05	-.08		
9. ED	-.02	.05	-.10	-.30	.25	.04	-.25	-.11	
10. PRO	-.02	-.09	.11	.27	.29	-.10	.13	.29	-.31

N of cases: 1299

Note. Correlations greater than or equal to .07 are significant, $p < .05$. MAN = managerial supervisory experience; ASSESSOR = experience with assessment centers; PRO = affiliation with the profession under consideration.

The testing strategy was to first do a principal components factor analysis loading all nine characteristics on nine factors and performing an orthogonal rotation. In this situation the concern was not for reducing the number of variables, but it was to reduce multicollinearity. The resulting nine factors are not correlated with each other. Table 4 displays the nine factors and the factor loadings of each of the assessor characteristics. Note that the values for each of the nine assessor characteristics were converted to Z scores for this analysis.

Table 4

Factor Loadings (Rotated Factor Matrix):

	Race (FAC 1)	Assessor Experience (FAC 2)	Gender (FAC 3)	Tenure (FAC 4)	Age (FAC 5)	Education (FAC 6)	Rank (FAC 7)	Pro (FAC 8)	Man (FAC 9)
RACE	.97	.01	.06	-.05	.00	-.14	.08	.11	.06
ASSESSOR	.01	.99	-.04	-.02	.03	-.04	-.01	.14	-.04
GENDER	.07	-.04	.96	.07	.08	-.04	.13	.04	.22
TENURE	-.05	-.02	.07	.95	.25	.02	.01	-.05	.17
AGE	.01	.04	.08	.25	.95	.01	.16	-.06	.06
ED	-.15	-.05	-.04	.02	.01	.95	.16	-.16	-.12
RANK	.09	-.01	.13	.02	.16	.16	.94	.17	.05
PRO	.12	.16	.04	-.05	-.06	-.16	.16	.95	.06
MAN	.06	-.05	.23	.18	.06	-.13	.05	.06	.94

Since this was not an experiment, but was a situation requiring use of existing data, multiple regression was the appropriate test. Because of the magnitude and extent of the intercorrelations among the independent variables, the uncorrelated factors were the most appropriate independent variables for use in the regression.

Preliminary tests were conducted to determine if the assumptions of the regression model were met. Plots showed that the distribution of residuals was approximately normal, and that the assumption of equal variance was met.

The next step in the process was to regress the assessor scores on the nine factors to determine which, if any, of the factors had a significant and important effect on the scores. The results of the regression are displayed in Table 5.

Table 5

Multiple Regression: Score regressed on Factors 1 through 9

Multiple R	.138
R Square	.019
Adjusted R Square	.017
Standard Error	12.36

Analysis of Variance

	df	Sum of Squares	Mean Square
Regression	2	3823.07	1911.53
Residual	1296	198033.07	152.80

F = 12.51**

Variables in the Equation

Variable	B	SE B	95% Confdnce Intrvl B		Beta	t
Factor 5(Age)	1.52	.3431	.848	2.194	.122	4.433**
Factor 5(Rank)	.80	.3431	.122	1.468	.064	2.317*
(Constant)	81.62	.3429	80.947	82.293		237.976**

* p <.03 ** p <.001

Discussion

Only two factors, AGE and RANK/position contributed significantly to the explained variance in SCOREs. The significant effect of management level measured in this study by RANK/position is not consistent with Carrier, et al. (1990). The lack of significant effect of selection experience or assessor training, measured in this study by ASSESSor experience, is consistent with Carrier, et al., (1990), and with Gaugler, et al., (1987).

Although the effects of AGE and RANK are significant, the adjusted R^2 value is less than .02, indicating that less than 2% of the variance in assessor scores is explained by these characteristics. There is a great deal of variance in scores left unexplained after the effect of assessor characteristics is removed. This result is not unexpected. Since the primary purpose of the assessment center process is to attempt to score assessees based on their actual performance in simulations, extraneous factors such as

assessor characteristics should have a minor effect, if any, on the final scores.

The only surprise here is how little the assessor characteristics affected the scores. The emphasis given to assessor training by the Guidelines, as well as the requirement that age, gender, ethnicity, and knowledge of the job should be considered suggested that these characteristics should have a more profound effect than was found. Of the issues raised by the Guidelines, only age was significant, but it was unimportant in terms of explained variance in scores.

Neither the suggestion by Thornton and Byham (1982) concerning assessor experience in the job (p.223), nor the Guideline's similar reference (Task Force, 1989, p. 462) is supported by these results. This is possibly one of the most important findings for those organizations that use nonaffiliated assessors.

The results of this study indicate that there are other factors at work to create the score. Hopefully the primary factor at work was the judgment of the assessors and was based only on the actual performance of the assessee. Certainly in this test, assessor characteristics were minor factors in the overall scoring paradigm.

Conclusions

Based on these analyses, I conclude that several forces were probably at work to keep the assessor scores so independent of assessor characteristics. These forces included: 1) the model used for selecting the assessors, 2) the training provided for the assessors, 3) the use of highly exercise specific behaviorally anchored rating forms, 4) and the use of the described integrating discussion technique after each exercise in eight of the nine centers.

The assessor selection model in the Appendix plainly pointed to using only experienced managers as assessors. The assessor characteristics in Table 2 show that the depth of experience of the assessors was such that one could expect them to be able to observe and evaluate relevant behaviors of managerial/supervisory level assessees.

The training provided for the assessors was both detailed and highly specific for each exercise. The assessors were able to become completely familiar with all aspects of each exercise, to include acting as both an assessee and assessor in one or more exercises.

The behaviorally anchored rating form used in every center was developed to give clear, unambiguous guidance to the assessors on what to look for in each exercise. The training given each assessor included the use of the form. The assessors were given the opportunity in the training session to discuss and ultimately come to an agreement with each other on what behaviors were considered appropriate or inappropriate.

Finally, in all but one center (Center 1), a non-evaluative integrating discussion was conducted after each exercise. (In Center 1 the discussion was conducted after all exercises were completed.) This procedure made it easier for the assessors to recall and relate behaviors they may not have had the opportunity to record during the exercise. It is difficult, if not impossible for an assessor to record all behaviors, especially if they are observing more than one assessee in a fast moving exercise. Karl & Wexley (1989) found that the accuracy of overall ratings was greatest when ratings collected after each exercise were averaged (p.5).

The requirement that no evaluative comments were to be made, or that scores were not to be divulged made it easier for each assessor to evaluate reported and observed behaviors against the global standard they brought to the center. Lowry (1991) found that using this procedure reduced interassessor influence.

It might be postulated that assessors who are not intimately familiar with the job under test would benefit from a full and free discussion including evaluative comments. In the absence of such a discussion, the argument might be made that nonaffiliated assessors' scores would probably vary widely from assessors who are more familiar with the job. In this study there was no significant difference among assessors who had experience in the job vs. those who did not. Hence, that proposition is not supported.

The assessor is clearly the single most important element in an assessment center. Therefore, they must selected with care. Assuming the assessors are selected, and an assessment center is conducted as described here, those individual characteristics of assessors studied here probably will not be an important factor in the scoring process.

These results support the Guidelines with respect to the need for training assessors. Assessor training was not directly measured in this study. Assessment center experience was measured, and was not a significant factor. The training provided all assessors was detailed, thorough, and the same in each center. The content of the training was in accord with the Guidelines (Task Force, 1989, pp. 465-466.) This special attention to training may well have been a major contributor to the lack of difference between the scores of the affiliated and nonaffiliated assessors.

The Guidelines are not specific concerning why gender, ethnicity, and age should be considered. One can imply that there is concern that these factors might affect scores. The results do not support that implication. However, ours is a multicultural society. It is important that those who judge candidates for important positions in the work force also reflect this diversity. Not because their judgments might be affected by those characteristics, but because such diversity among assessors brings a perspective to the process that is reflected in more subtle ways.

The nine assessment centers that were examined reflect a sizable, yet not completely representative sample of assessment centers

conducted in the public sector. I do not suggest, therefore, that these results be generalized completely to all assessment centers. I do suggest that these results provide reasonable support for the following proposition:

Assessor characteristics will not have a major affect on scores if certain rules for the selection of assessors and conduct of the assessment center are followed. These rules are: 1) assessors should be selected from among successful managers using the guidelines described in the Appendix; 2) assessor training should be specific for each assessment center and should generally include the content described in the Guidelines (Task Force, 1989, pp. 465-466); 3) behaviorally anchored observation and rating forms should be used for each exercise; 4) non-evaluative integrating discussions should be used after each exercise, and no scores should be divulged to other assessors (Lowry, 1991); and 5) despite the lack of major effects of gender, sex, and ethnicity, assessor selection should consider these factors.

Researchers and practitioners should carefully examine these rules to determine whether they are appropriate in a wider variety of situations.

Appendix

Criteria for Selecting Assessors [1]

The assessors:

1) should not know the candidates, except in a casual way.

2) must be able to observe and evaluate relevant behaviors.

3) should be willing to spend extra time in the evaluation process.

4) must be perceived by the candidates as qualified to evaluate their behavior.

5) should be knowledgeable of the tasks to be performed and what is expected of personnel in the position being evaluated.

6) should have had sufficient experience in supervision and management in equivalent organizations or positions to be able to observe, classify, and evaluate relevant behaviors.

7) should have formal education/technical training at the same or higher level as the average incumbent in the position under consideration.

8) should be capable of being trained to ensure that they can effectively and efficiently perform their function in the process.

9) should have participated in other assessment centers or have received prior training in the process.

10) generally should come from equivalent organizations and be in the same or higher level positions as the job under consideration.

11) ethnic and gender diversity should be considered.

The number of nonaffiliated assessors should generally not exceed 50% of the total. The preferred ratio is one nonaffiliated for three affiliated assessors. It is considered essential that the nonaffiliated assessors be made aware of the technical complexities of the jobs under consideration, and that they be told they will evaluate how well the assessee performs simulated tasks that include complex technical issues. Select only those nonaffiliated assessors who meet the above criteria.

Note [1]: These criteria are based in part on the Guidelines (Task Force,1989).

References

Bycio, P., Alvares, K. M., & Hahn, J. (1987). Situational specificity in assessment center ratings: A confirmatory factor analysis. Journal of Applied Psychology, 72, 463-474.

Carrier, M. R., Dalessio, A. T., & Brown, S. H. (1990). Correspondence between estimates of content and criterion-related validity values. Personnel Psychology, 43, 85-100.

Dugan, B. (1988). Effects of assessor training on information use. Journal of Applied Psychology, 73, 743-748.

Fitzgerald, L.F., & Quaintance, M. K. (1982). Survey of assessment center use in state and local government. Journal of Assessment Center Technology, 5, 9-21.

Karl, K. A., & Wexley, K. N. (1989). Patterns of performance and rating frequency: Influence on the assessment of performance. Journal of Management, 15, 5-20.

Lowry, P. E. (1991). The Assessment center: Reducing interassessor influence. Public Personnel Management, 20, 19-26.

Reilly, R.R., Henry, S., & Smither, J.W. (1990). An examination of the effects of using behavior checklists on the construct validity of assessment center dimensions. Personnel Psychology, 43, 71-84.

Robertson, I., Gratton, L., & Sharpley, D. (1987). The psychometric properties and design of managerial assessment centres: Dimensions into exercises won't go. Journal of Occupational Psychology, 60, 187-195.

Russell, C. J. (1985). Individual decision processes in an assessment center. Journal of Applied Psychology, 70, 737-746.

Sackett, P. R., & Wilson, M. A. (1982). Factors affecting the consensus judgment process in managerial assessment centers. Journal of Applied Psychology, 67, 10-17.

Shore, T. H., Thornton, III, G. C., & Shore, L. M. (1990). Construct validity of two categories of assessment center dimension ratings. Personnel Psychology, 43, 101-116.

Task Force on Assessment Center Guidelines (1989). Guidelines and ethical considerations for assessment center operations. Public Personnel Management, 18, 457-470.

Task Force on Assessment Center Standards (1980). Standards and ethical considerations for assessment center operations. The Personnel Administrator, 35-38.

Thornton, III, G. C. & Byham, W. C. (1982). Assessment centers and managerial performance. New York: Academic Press.

Yeager, S. J. (1982). Use of assessment centers by metropolitan fire departments in North America. Public Personnel Management, 15, 51-64.